



Likelihood Inference for Exponential-Trawl Processes

Citation

Shephard, Neil, and Justin J. Yang. 2016. "Likelihood inference for exponential-trawl processes." In *The Fascination of Probability, Statistics and their Applications*, eds. Podolskij, M., Stelzer, R., Thorbjørnsen, S., Veraart, A.E.D.: 251-281. Cham, Switzerland: Springer International.

Published Version

doi:10.1007/978-3-319-25826-3_12

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:34650467>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Likelihood Inference for Exponential-Trawl Processes

Neil Shephard and Justin J. Yang

Abstract Integer-valued trawl processes are a class of serially correlated, stationary and infinitely divisible processes that Ole E. Barndorff-Nielsen has been working on in recent years. In this Chapter, we provide the first analysis of likelihood inference for trawl processes by focusing on the so-called exponential-trawl process, which is also a continuous time hidden Markov process with countable state space. The core ideas include prediction decomposition, filtering and smoothing, complete-data analysis and EM algorithm. These can be easily scaled up to adapt to more general trawl processes but with increasing computation efforts.

1 Introduction

In recent years, Ole E. Barndorff-Nielsen has been working on a class of stochastic models called integer-valued trawl processes. References include [2], [4] and [5]. These are flexible models whose core randomness is driven by Poisson random measures. Trawl processes are related to the up-stairs processes of [25] and the random measure processes of [24]. Both of these processes are stationary. [5] also brings out the relationship between their processes and $M/G/\infty$ queues (e.g. [16], [18] and [8, Ch. 6.31]) and mixed moving average processes (e.g. [22]). Related discrete time count models include [9], [13], [10], [12], [15], [17], [26], [14], [17] and [23]. Trawl processes also fall within the wide class of the so-called ambit fields

Neil Shephard
Department of Economics, Harvard University
1802 Cambridge Street, 02138, Cambridge, MA, USA
e-mail: shephard@fas.harvard.edu

Justin J. Yang
Department of Statistics, Harvard University
1 Oxford Street, 02138, Cambridge, MA, USA
e-mail: juchenjustinyang@fas.harvard.edu

(e.g. [7] and [3]). Recently, [21] models high frequency financial data by using a trawl process to allow for fleeting movements to prices in addition to an integer-valued Lévy process proposed by [6].

As far as we know, there is no existing literature that directly and completely addresses likelihood inference for these trawl processes—or equivalently the prediction based upon it. Even though there are a large number of papers that focus on likelihood inference for marked point processes (see [11] for a survey), it only indirectly and partially describes trawl processes in terms of their jumps. A thorough likelihood inference for trawl processes needs to include the information in the initial value of the process.

In this Chapter, we provide a thorough analysis of likelihood inference for integer-valued trawl processes and demonstrate the core ideas—prediction decomposition, filtering, smoothing and EM algorithm—by focusing on the so-called exponential trawl. It is not only a simplification of the modelling framework but also an intellectually interesting special case of its own, as in this special case the resulting trawl process is a continuous time hidden Markov process with countable state space. The theoretical analysis for the filtering and smoothing problems for this type of process has been discussed in details by [19] and [20], using the classical theory of Kolmogorov’s forward and backward differential equations. We particularly emphasize that the resulting EM algorithm in this special case is exact in the sense that there are no discretization errors in its computation.

The major goal of this Chapter is to derive filtering and smoothing results in the framework of trawl processes, so the analysis adopted here can be easily scaled up to adapt to the discussions of other general trawls or even the inclusion of a non-stationary component proposed in [21]. These general discussions will be dealt with elsewhere, for they require a significantly more sophisticated particle filtering and smoothing device. We also discuss non-negative trawl processes, which are particularly easy to work with.

The structure of this Chapter is as follows. In Section 2, we remind the reader how to construct trawl processes using the exponential trawl. Section 3 includes details of how to carry out filtering and smoothing for these models. In Section 4, we show likelihood inference for exponential-trawl processes based on these filters and smoothers. Section 5 discusses the important but analytically tractable special case of non-negative trawl processes. We finally conclude in Section 6. The Appendix contains the proofs and derivations of various results given in this Chapter.

2 Exponential-Trawl Processes

In this Section, we build our notation, definitions and key structures for the exponential-trawl process that will be focused on throughout this Chapter. We also provide its log-likelihood function based on observed data.

2.1 Definition

Our model will be based on a homogeneous Lévy basis on $[0, 1] \times \mathbb{R} \mapsto \mathbb{Z} \setminus \{0\}$, which models the discretely scattered events of integer size (with direction) $y \in \mathbb{Z} \setminus \{0\}$ at each point in time $s \in \mathbb{R}$ and height $x \in [0, 1]$. It is defined by

$$L(dx, ds) \triangleq \int_{-\infty}^{\infty} y N(dy, dx, ds), \quad (x, s) \in [0, 1] \times \mathbb{R},$$

where N is a three-dimensional Poisson random measure with intensity measure

$$\mathbb{E}(N(dy, dx, ds)) = \nu(dy) dx ds.$$

Here ds means the arrival times are uniformly scattered (over \mathbb{R}), dx means the random heights are also uniformly scattered (over $[0, 1]$) and $\nu(dy)$ is a Lévy measure concentrated on the non-zero integers $\mathbb{Z} \setminus \{0\}$. Without any confusion, we will abuse the notation $\nu(y)$ to denote the mass of the Lévy measure centered at y . Throughout this Chapter, we assume that

$$\int_{-\infty}^{\infty} \nu(dy) = \sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y) < \infty.$$

Following [5], we think of dragging a fixed Borel measurable set $A \subseteq [0, 1] \times (-\infty, 0]$ through time

$$A_t \triangleq A + (0, t), \quad t \geq 0,$$

so the trawl process is defined by

$$Y_t \triangleq L(A_t) = \int_{[0, 1] \times \mathbb{R}} 1_A(x, s - t) L(dx, ds).$$

Throughout the rest of this Chapter, we will focus on the exponential trawl

$$A \triangleq \{(x, s) : s \leq 0, 0 \leq x < d(s) \triangleq \exp(s\phi)\}, \quad \phi > 0,$$

to simplify our exposition of the key ideas. We will leave results on more general trawls in another study.

Example 1. Suppose that

$$\nu(dy) = \nu^+ \delta_{\{1\}}(dy) + \nu^- \delta_{\{-1\}}(dy), \quad \nu^+, \nu^- > 0,$$

where $\delta_{\{\pm 1\}}(dy)$ is the Dirac point mass measure centered at ± 1 . The corresponding $L(dx, ds)$ is called a Skellam Lévy basis, while the special case of $\nu^- = 0$ is called Poisson. The upper panel of Fig. 1 shows events in L using $\nu^+ = \nu^- = 10$, taking sizes on $1, -1$ with black and white dots respectively and with equal probability. The lower panel of Fig. 1 then illustrates the resulting Skellam exponential-trawl process $Y_t = L(A_t)$ using $\phi = 2$, which sums up all the effects (both positive

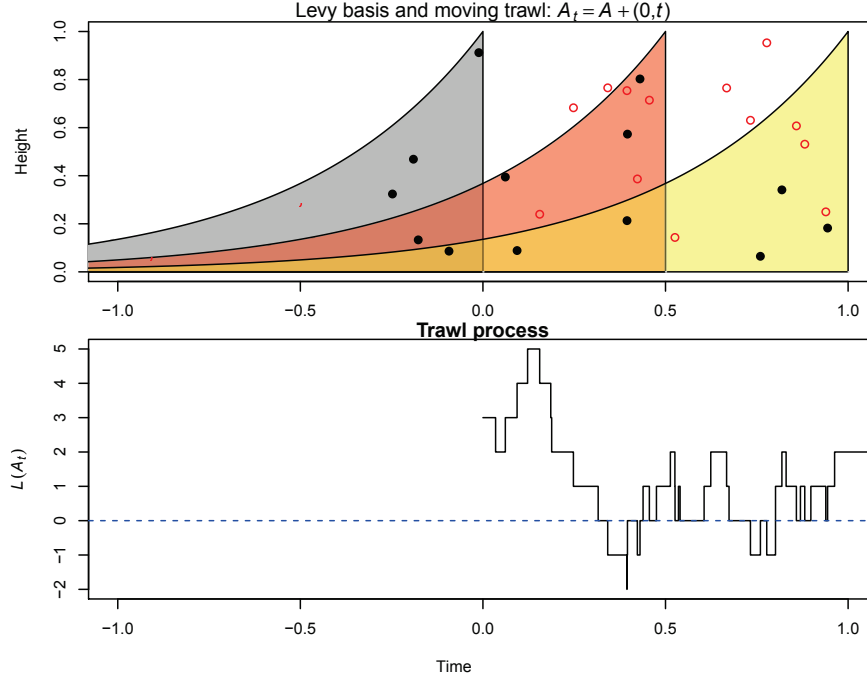


Fig. 1 A moving trawl A_t is joined by the Skellam Lévy basis $L(dx, ds)$, where the horizontal axis s is time and the vertical axis x is height. The shaded area is an example of the exponential trawl A , while we also show the outlines of A_t when $t = 1/2$ and $t = 1$. Also shown below is the implied trawl process $Y_t = L(A_t)$. Code: `EPTprocess.Illustration.R`

and negative) captured by the exponential trawl. Dynamically, $L(A_t)$ will move up by 1 if the moving trawl A_t either captures one positive event or releases a negative one; conversely, it will move down by 1 if vice versa. Notice that $Y_0 = L(A_0)$ might not be necessarily zero and the path of Y at negative time is not observed.

2.2 Markovian Counting Process

For $y \in \mathbb{Z} \setminus \{0\}$, let $C_t^{(y)} \in \{0, 1, 2, \dots\}$ be the total counts of surviving events of size y in the trawl at time t , which also includes the event that arrives *exactly* at time t , so each $C_t^{(y)}$ must be càdlàg (right-continuous with left-limits). Then clearly the trawl process can be represented as

$$Y_t = \sum_{y \in \mathbb{Z} \setminus \{0\}} y C_t^{(y)}, \quad t \geq 0. \quad (1)$$

Note that each $C_t^{(y)}$ is not only a Poisson exponential-trawl process with (different) intensity of arrivals $\nu(y)$ (and sharing the same trawl) but also a $M/M/\infty$ queue and hence a continuous time Markov process. Hence, for $\{\mathcal{C}_t^{(y)}\}_{t \geq 0}$ being the natural filtration generated by the counting process $C_t^{(y)}$, i.e., $\mathcal{C}_t^{(y)} \triangleq \sigma\left(\{C_s^{(y)}\}_{0 \leq s \leq t}\right)$, it has (infinitesimal) transition probabilities (or rates or intensities)

$$\lim_{dt \rightarrow 0} \frac{\mathbb{P}\left(C_t^{(y)} - C_{t-dt}^{(y)} = j \mid \mathcal{C}_{t-dt}^{(y)}\right)}{dt} = \begin{cases} \nu(y), & \text{if } j = 1 \\ \phi C_{t-}^{(y)}, & \text{if } j = -1 \\ 0, & \text{if } j \in \mathbb{Z} \setminus \{-1, 1\} \end{cases}. \quad (2)$$

The cases of $j = 1$ or -1 —which correspond to the arrival of a new event of size y and the departure of an old one—are the only two possible infinitesimal movements of $C_t^{(y)}$ due to the point process nature of the Lévy basis. Note that the arrival rate and departure rate are controlled by the Lévy measure ν and the trawl parameter ϕ respectively. Derivation of (2) can be found in many standard references for queue theory (e.g. [1]).

Remark 1. Let $\Delta X_t \triangleq X_t - X_{t-}$ denote the *instantaneous* jump of any process X at time t . Then the transition probability (2) can be conveniently written in a differential form

$$\mathbb{P}\left(\Delta C_t^{(y)} = j \mid \mathcal{C}_{t-}^{(y)}\right) = \begin{cases} \nu(y) dt, & \text{if } j = 1 \\ \phi C_{t-}^{(y)} dt, & \text{if } j = -1 \\ 0, & \text{if } j \in \mathbb{Z} \setminus \{-1, 1\} \end{cases}.$$

Throughout this Chapter, our analysis will be majorly based on this infinitesimal point of view for the ease of demonstration. All of our arguments can be rephrased in a mathematically tighter way.

The independence property of the Lévy basis implies the independence between each $C_t^{(y)}$ for $y \in \mathbb{Z} \setminus \{0\}$, so the joint count process

$$\mathbf{C}_t \triangleq \left(\dots, C_t^{(-2)}, C_t^{(-1)}, C_t^{(1)}, C_t^{(2)}, \dots \right)$$

is also Markovian, which serves as the unobserved state process for the observed *hidden Markov process* Y_t and will be the central target for the filter and smoother we will discuss in a moment. Let $\mathcal{C}_t \triangleq \sigma(\{\mathbf{C}_s\}_{0 \leq s \leq t}) = \bigvee_{y \in \mathbb{Z} \setminus \{0\}} \mathcal{C}_t^{(y)}$ be the join filtration. Clearly, from (2), \mathbf{C}_t has (infinitesimal) transition probabilities

$$\mathbb{P}(\Delta \mathbf{C}_t = \mathbf{j} \mid \mathcal{C}_{t-}) = \begin{cases} \nu(y) dt, & \text{if } \mathbf{j} = \mathbf{1}^{(y)} \text{ for some } y \\ \phi C_{t-}^{(y)} dt, & \text{if } \mathbf{j} = -\mathbf{1}^{(y)} \text{ for some } y, \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where $\mathbf{1}^{(y)} \in \mathbb{Z}^\infty$ is the vector that takes 1 at y -th component and 0 otherwise.

The trawl process Y_t can be also written as

$$Y_t = \sum_{y=1}^{\infty} y Y_t^{(y)}, \quad Y_t^{(y)} \triangleq C_t^{(y)} - C_t^{(-y)},$$

where each $Y_t^{(y)}$ is a Skellam exponential-trawl process. Each $Y_t^{(y)}$ is observed from the path of Y_t up to its initial value $Y_0^{(y)}$, for we can exactly observe all the jumps of Y_t and hence allocate them into the appropriate $Y_t^{(y)}$. In other words, we can regard the observed trawl process as (i) a *marked point process* $\Delta Y_t \in \mathbb{Z} \setminus \{0\}$, which consists of several independent (given all the $Y_0^{(y)}$) marked point process $\Delta Y_t^{(y)} \in \{-1, 1\}$, plus (ii) the initial value Y_0 . The missing components $Y_0^{(y)}$'s will have some mild effects on $\Delta Y_t^{(y)}$. It is this initial value challenge that differentiates the likelihood analysis of trawl processes from that of marked point processes.

The special case where Y_t is always non-negative has further simpler structure, as we must have $C_t^{(-y)} = 0$ for all $y = 1, 2, \dots$ and hence $C_t^{(y)} = Y_t^{(y)}$ is directly observed up to its initial condition $C_0^{(y)}$, which can be well-approximated if the observation period T is large enough. We will go through these details in Section 5.

2.3 Conditional Intensities and Log-likelihood

Let $\{\mathcal{F}_t\}_{t \geq 0}$ be the natural filtration generated by the observed trawl process Y_t , i.e. $\mathcal{F}_t \triangleq \sigma(\{Y_s\}_{0 \leq s \leq t})$. Define the càdlàg conditional intensity process of the trawl process Y as

$$\lambda_{t-}^{(y)} \triangleq \lim_{dt \rightarrow 0} \frac{\mathbb{P}(Y_t - Y_{t-dt} = y | \mathcal{F}_{t-dt})}{dt}, \quad y \in \mathbb{Z} \setminus \{0\}, t > 0 \quad (4)$$

or conveniently in a differential form

$$\lambda_{t-}^{(y)} dt \triangleq \mathbb{P}(\Delta Y_t = y | \mathcal{F}_{t-}). \quad (5)$$

It means the (time-varying) predictive intensity of a size y move at time t of the trawl process, conditional on information instantaneously before time t .

Remark 2. To emphasize the \mathcal{F}_t -predictability of $\lambda^{(y)}$, i.e., being adapted to the left natural filtration \mathcal{F}_{t-} , we will keep the subscript $t-$ throughout this Chapter. This is particularly informative in the implementation of likelihood calculations, reminding us to take the *left-limit* of the intensity process whenever there is a jump.

For any two σ -fields \mathcal{F} and \mathcal{G} , let the Radon-Nikodym derivative over $\mathcal{F}|\mathcal{G}$ between two probability measures \mathbb{P} and \mathbb{Q} be

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}|\mathcal{G}} \triangleq \frac{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F} \vee \mathcal{G}}}{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{G}}}.$$

In particular, when $\mathcal{G} = \sigma(X)$ for any random variable X , we will simply write the subscript as $\mathcal{F}|X$. The following classical result serves as the foundation for all likelihood inference for jump processes.

Theorem 1. *Let X_t be any integer-valued stochastic process and $\{\mathcal{F}_t^X\}_{t \geq 0}$ be its associated natural filtration. Assume that, under both \mathbb{P} and \mathbb{Q} , (i) it has finite expected number of jumps during $(0, T]$, and (ii) the conditional intensities $\lambda_{t-}^{(y), \mathbb{P}}$ and $\lambda_{t-}^{(y), \mathbb{Q}}$ are well-defined using (4) and \mathcal{F}_t^X . Then $\mathbb{P} \ll \mathbb{Q}$ over $\mathcal{F}_T^X|X_0$ if and only if $\lambda_{t-}^{(y), \mathbb{Q}}$ is strictly positive. In this case, the logarithmic Radon-Nikodym derivative over $\mathcal{F}_T^X|X_0$ is*

$$\begin{aligned} \log \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_T^X|X_0} &= \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \left(\frac{\lambda_{t-}^{(y), \mathbb{P}}}{\lambda_{t-}^{(y), \mathbb{Q}}}\right) 1_{\{\Delta X_t = y\}} \\ &\quad - \int_{t \in (0, T]} \sum_{y \in \mathbb{Z} \setminus \{0\}} \left(\lambda_{t-}^{(y), \mathbb{P}} - \lambda_{t-}^{(y), \mathbb{Q}}\right) dt. \end{aligned}$$

Proposition 14.4.I of [11] provides a complete and mathematically rigorous treatment for this Theorem. For completeness, we also provide an intuitive and heuristic derivation in the Appendix. A direct application of Theorem 1 gives the following Corollary.

Corollary 1. *The log-likelihood function of the (general) trawl process is (ignoring the constant)*

$$l_{\mathcal{F}_T}(\theta) = \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \lambda_{t-}^{(y)} 1_{\{\Delta Y_t = y\}} - \int_{t \in (0, T]} \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y)} dt + l_{Y_0}(\theta), \quad (6)$$

where the parameters of interest θ include the Lévy measure $\nu(dy)$ (i.e. $\nu(y)$'s) and the trawl parameter ϕ .

The study of likelihood inference for trawl processes then reduces to the calculations of conditional intensities $\lambda_{t-}^{(y)}$ for $y \in \mathbb{Z} \setminus \{0\}$. Now, by law of iterated expectations and the fact that $\mathcal{C}_t \supseteq \mathcal{F}_t$ for all t (because of (1)), we have

$$\begin{aligned} \lambda_{t-}^{(y)} dt &= \mathbb{E}(\mathbb{P}(\Delta Y_t = y | \mathcal{C}_{t-}) | \mathcal{F}_{t-}) \\ &= \mathbb{E}\left(\mathbb{P}\left(\Delta \mathbf{C}_t = \mathbf{1}^{(y)} \middle| \mathcal{C}_{t-}\right) \middle| \mathcal{F}_{t-}\right) + \mathbb{E}\left(\mathbb{P}\left(\Delta \mathbf{C}_t = -\mathbf{1}^{(-y)} \middle| \mathcal{C}_{t-}\right) \middle| \mathcal{F}_{t-}\right) \\ &= \nu(y) dt + \phi \mathbb{E}\left(C_{t-}^{(-y)} \middle| \mathcal{F}_{t-}\right) dt, \end{aligned}$$

where the second line follows because the event $\Delta Y_t = y$ must come from either an arrival of a new size y event or a departure of an old size $-y$ event; the third line follows from (3). Thus,

$$\lambda_{t-}^{(y)} = \nu(y) + \phi \mathbb{E} \left(C_{t-}^{(-y)} \middle| \mathcal{F}_{t-} \right), \quad y \in \mathbb{Z} \setminus \{0\}. \quad (7)$$

In next Section, we will study an exact filtering scheme to numerically calculate $\mathbb{E} \left(C_{t-}^{(-y)} \middle| \mathcal{F}_{t-} \right)$.

The non-negative exponential-trawl process, where we always have positive events, admits a further simplification

$$\lambda_{t-}^{(y)} = \nu(y), \quad \lambda_{t-}^{(-y)} = \phi \mathbb{E} \left(C_{t-}^{(y)} \middle| \mathcal{F}_{t-} \right), \quad y = 1, 2, \dots, \quad (8)$$

so likelihood inference for such a case is easier. In the Poisson case, all the impacts are of size one, so in particular $C_0^{(1)} = Y_0$ is also observed (as $C_0^{(y)} = 0$ for all $y \neq 1$), which allows us to bypass the conditional expectation in (8) for $y = 1$.

3 Exact Filter and Smoother for Exponential-Trawl Processes

3.1 Filtering

In general we need to solve the filtering problems for \mathbf{C}_t to implement (6) and (7). Denote the filtering probability mass function as

$$p_{t,s}(\mathbf{j}) \triangleq \mathbb{P}(\mathbf{C}_t = \mathbf{j} \mid \mathcal{F}_s), \quad \mathbf{j} = (\dots, j_{-2}, j_{-1}, j_1, j_2, \dots), \quad j_y = 0, 1, 2, \dots, \quad t, s \geq 0.$$

Also, let $\|\mathbf{j}\|_1 \triangleq \sum_{y \in \mathbb{Z} \setminus \{0\}} j_y$ and $D_t \triangleq \|\mathbf{C}_t\|_1 = \sum_{y \in \mathbb{Z} \setminus \{0\}} C_t^{(y)}$.

Our goal here is to sequentially update $p_{t-,t-}(\mathbf{j})$, where the initial distribution is derived from

$$C_0^{(y)} \overset{\text{indep.}}{\sim} \text{Poisson} \left(\frac{\nu(y)}{\phi} \right) \text{ subject to } \sum_{y \in \mathbb{Z} \setminus \{0\}} y C_0^{(y)} = Y_0,$$

so, by letting $\text{Poisson}(x|\lambda) \triangleq \lambda^x e^{-\lambda} / x!$, we have

$$p_{0,0}(\mathbf{j}) = \frac{\prod_{y \in \mathbb{Z} \setminus \{0\}} \text{Poisson}(j_y | \nu(y) / \phi)}{\mathbb{P} \left(\sum_{y \in \mathbb{Z} \setminus \{0\}} y C_0^{(y)} = Y_0 \right)},$$

where the denominator can be numerically calculated using the inverse fast Fourier transform [21].

Notice that the filtering distribution not only updates at the times when the process jumps but also at those inactivity periods. We discuss these two cases separately.

Theorem 2 (Forward Filtering).

1. *[Update by inactivity]* Assume that the last jump time is τ (or $\tau = 0$) and the current time is $t-$, where $\Delta Y_s = 0$ for $\tau < s < t$ (and $\Delta Y_\tau \neq 0$ if $\tau > 0$). Then

$$p_{t-,t-}(\mathbf{j}) = \frac{e^{-\phi \|\mathbf{j}\|_1 (t-\tau)} p_{\tau,\tau}(\mathbf{j})}{\sum_{\mathbf{k}} e^{-\phi \|\mathbf{k}\|_1 (t-\tau)} p_{\tau,\tau}(\mathbf{k})}, \quad (9)$$

where $p_{\tau,\tau}$ is the filtering distribution we have already known at time τ .

2. *[Update by jump]* Assume that the current time is $\tau-$ and $\Delta Y_\tau = y$ for some $y \in \mathbb{Z} \setminus \{0\}$. Then

$$p_{\tau,\tau}(\mathbf{j}) = \frac{1}{\lambda_{\tau-}^{(y)}} \left(v^{(y)} p_{\tau-,\tau-}(\mathbf{j} - \mathbf{1}^{(y)}) + \phi (j_{-y} + 1) p_{\tau-,\tau-}(\mathbf{j} + \mathbf{1}^{(-y)}) \right), \quad (10)$$

where $p_{\tau-,\tau-}$ is the filtering distribution we have already known at time $\tau-$.

Overall, the filtering procedures (9) and (10) imply that $p_{t-,t-}(\mathbf{j})$ can be updated in continuous time without discretization errors at any set of finite discrete time points, so we call it an exact filter.

Example 2. For Skellam exponential-trawl process with Lévy intensities v^+ and v^- , we always have

$$Y_{t-} = C_{t-}^{(+)} - C_{t-}^{(-)}, \quad t > 0,$$

so knowing $p_{t-,t-}(j) \triangleq \mathbb{P}\left(C_{t-}^{(-)} = j \mid \mathcal{F}_{t-}\right)$ immediately gives us $p_{t-,t-}(j, k)$. Hence, the filtering updating scheme reduces to the following: starting from $\tau = 0$,

$$\begin{aligned} p_{t-,t-}(j) &\propto e^{-\phi(2j+Y_\tau)(t-\tau)} p_{\tau,\tau}(j) \quad \text{if } \Delta Y_s = 0 \text{ for } \tau < s < t, \\ p_{\tau,\tau}(j) &\propto v^+ p_{\tau-,\tau-}(j) + \phi(j+1) p_{\tau-,\tau-}(j+1) \quad \text{if } \Delta Y_\tau = 1, \\ p_{\tau,\tau}(j) &\propto v^- p_{\tau-,\tau-}(j-1) + \phi(j+Y_{\tau-}) p_{\tau-,\tau-}(j) \quad \text{if } \Delta Y_\tau = -1. \end{aligned}$$

We then renormalize $p_{t-,t-}(j)$ such that $\sum_{j=0}^{\infty} p_{t-,t-}(j) = 1$ in each step of the updates. Knowing the filtering distributions $p_{t-,t-}(j)$ allows us to calculate

$$\mathbb{E}\left(C_{t-}^{(-)} \mid \mathcal{F}_{t-}\right) = \sum_{j=0}^{\infty} j p_{t-,t-}(j), \quad \mathbb{E}\left(C_{t-}^{(+)} \mid \mathcal{F}_{t-}\right) = \sum_{j=0}^{\infty} j p_{t-,t-}(j) + Y_{t-}.$$

Using the following settings, with time unit being second,

$$v^+ = 0.013, \quad v^- = 0.011, \quad \phi = 0.034, \quad T = 21 \times 60^2 = 75,600 \text{ (sec.)}, \quad (11)$$

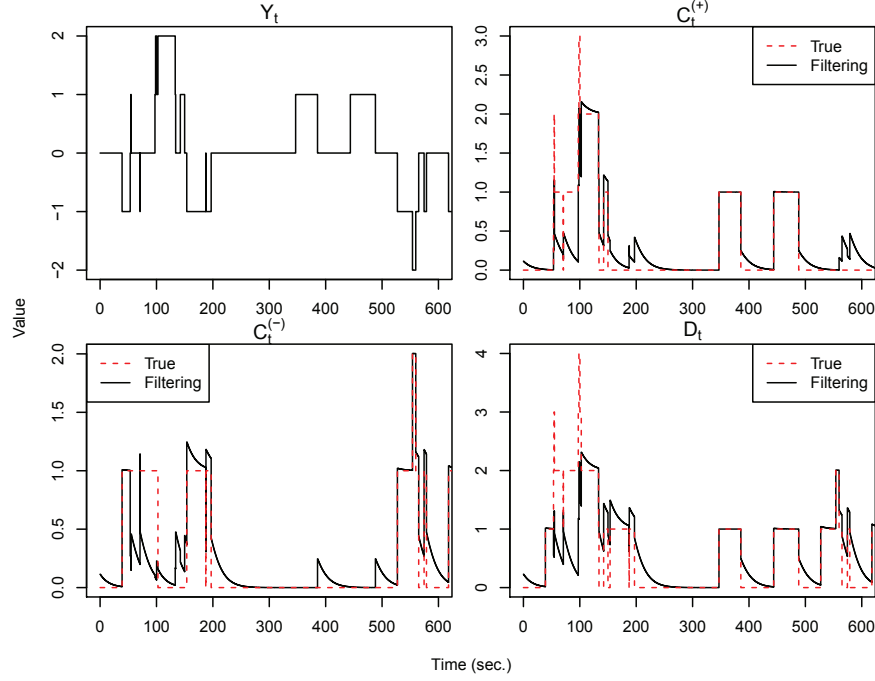


Fig. 2 *Top left:* A simulated path for the Skellam exponential-trawl process Y_t . *Top right, Bottom left, Bottom right:* Paths of the true hidden counting processes $C_t^{(+)}$, $C_t^{(-)}$ and $D_t = C_t^{(+)} + C_t^{(-)}$ of surviving events in the trawl along with their filtering estimations. Code: EPTprocess.FilteringSmoothing.Illustration.R

Fig. 2 shows a simulated path of the trawl process Y_t together with the filtering expectations of $C_t^{(+)}$, $C_t^{(-)}$ and $D_t = C_t^{(+)} + C_t^{(-)}$, the total number of surviving (both positive and negative) events in the trawl at time t .

3.2 Smoothing

We now consider the smoothing procedure for the exponential-trawl process Y_t , which is necessary for the likelihood inference based on the EM algorithm we will see in a moment.

Running the filtering procedure up to time T , we then start from $p_{T,T}$ to conduct the smoothing procedure.

Theorem 3 (Backward Smoothing).

1. **[Update by inactivity]** Assume that the (backward) last jump time is τ (or $\tau = T$) and the current time is t , where $\Delta Y_s = 0$ for $t \leq s < \tau$ (and $\Delta Y_\tau \neq 0$ if $\tau < T$). Then

$$p_{t,T}(\mathbf{j}) = p_{\tau-,T}(\mathbf{j}),$$

where $p_{\tau-,T}$ is the smoothing distribution we have already known at time $\tau-$.

2. **[Update by jump]** Assume that the current time is τ and $\Delta Y_\tau = y$ for some $y \in \mathbb{Z} \setminus \{0\}$. Then

$$p_{\tau-,T}(\mathbf{j}) = \frac{p_{\tau-, \tau-}(\mathbf{j})}{\lambda_{\tau-}^{(y)}} \left(v(y) \frac{p_{\tau,T}(\mathbf{j} + \mathbf{1}^{(y)})}{p_{\tau,\tau}(\mathbf{j} + \mathbf{1}^{(y)})} + \phi j_{-y} \frac{p_{\tau,T}(\mathbf{j} - \mathbf{1}^{(-y)})}{p_{\tau,\tau}(\mathbf{j} - \mathbf{1}^{(-y)})} \right), \quad (12)$$

where $p_{\tau-, \tau-}$ and $p_{\tau,\tau}$ are from the forward filtering procedure and $p_{\tau,T}$ is the smoothing distribution we have already known at time τ .

The two terms in (12) are

$$\mathbb{P}\left(\mathbf{C}_{\tau-} = \mathbf{j}, \mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)} \mid \mathcal{F}_T\right) \text{ and } \mathbb{P}\left(\mathbf{C}_{\tau-} = \mathbf{j}, \mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(y)} \mid \mathcal{F}_T\right)$$

respectively, so, in particular,

$$\mathbb{P}\left(\Delta C_\tau^{(y)} = 1 \mid \mathcal{F}_T\right) = \sum_{\mathbf{j}} \frac{p_{\tau-, \tau-}(\mathbf{j})}{\lambda_{\tau-}^{(y)}} \left(v(y) \frac{p_{\tau,T}(\mathbf{j} + \mathbf{1}^{(y)})}{p_{\tau,\tau}(\mathbf{j} + \mathbf{1}^{(y)})} \right), \quad (13)$$

$$\mathbb{P}\left(\Delta C_\tau^{(y)} = -1 \mid \mathcal{F}_T\right) = \sum_{\mathbf{j}} \frac{p_{\tau-, \tau-}(\mathbf{j})}{\lambda_{\tau-}^{(y)}} \left(\phi j_{-y} \frac{p_{\tau,T}(\mathbf{j} - \mathbf{1}^{(-y)})}{p_{\tau,\tau}(\mathbf{j} - \mathbf{1}^{(-y)})} \right). \quad (14)$$

These (total) weights in (12) will be recorded for every jump time τ as by-products of the smoothing procedure, for later they will play important roles in the EM algorithm introduced in Subsection 4.3.

Example 3 (Continued from Example 2).

For Skellam exponential-trawl process, the smoothing updating scheme reduces to the following: starting from $\tau = T$,

$$\begin{aligned} p_{t,T}(j) &= p_{\tau-,T}(j) \quad \text{if } \Delta Y_s = 0 \text{ for } t \leq s < \tau, \\ p_{\tau-,T}(j) &\propto p_{\tau-, \tau-}(j) \left(v^+ \frac{p_{\tau,T}(j)}{p_{\tau,\tau}(j)} + \phi j \frac{p_{\tau,T}(j-1)}{p_{\tau,\tau}(j-1)} \right) \quad \text{if } \Delta Y_\tau = 1, \\ p_{\tau-,T}(j) &\propto p_{\tau-, \tau-}(j) \left(v^- \frac{p_{\tau,T}(j+1)}{p_{\tau,\tau}(j+1)} + \phi (Y_{\tau-} + j) \frac{p_{\tau,T}(j)}{p_{\tau,\tau}(j)} \right) \quad \text{if } \Delta Y_\tau = -1. \end{aligned}$$

We also renormalize $p_{t,T}(j)$ in each step of the updates.

Using the same simulated path and the same setting (11) as in Example 2, we show the smoothing expectations of $C_t^{(+)}$, $C_t^{(-)}$ and D_t in Fig. 3. For most of the time, the smoothing expectations can match the truth quite well and will remove those peaks of filtering expectations resulted from departures (such as the one close to $t = 400$ in the plot for $C_t^{(-)}$).

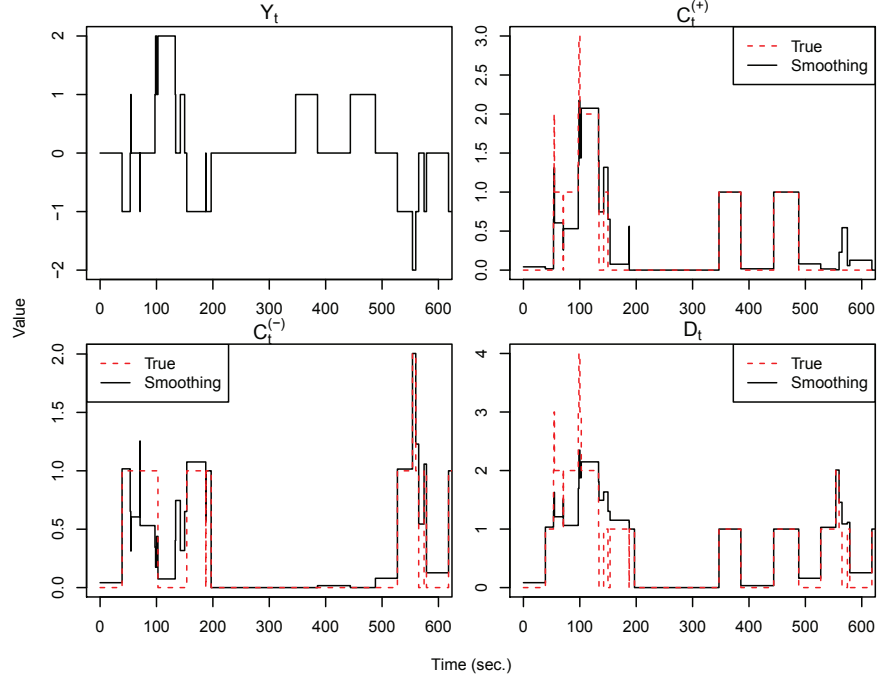


Fig. 3 *Top left:* A simulated path for the Skellam exponential-trawl process Y_t . *Top right, Bottom left, Bottom right:* Paths of the true hidden counting processes $C_t^{(+)}$, $C_t^{(-)}$ and $D_t = C_t^{(+)} + C_t^{(-)}$ of surviving events in the trawl along with their smoothing estimations. Code: EPTprocessFilteringSmoothingIllustration.R

Now we are capable of conducting likelihood inference for exponential-trawl processes as one of the most important applications of the filtering and smoothing procedures we have already built here.

4 Likelihood Inference for General Exponential-Trawl Processes

It has been reported by [5] and [21] that the moment-based inference for the family of trawl processes could be easily performed, but such inference is arbitrarily dependent on its procedure design. In this Section, we focus on the maximum likelihood estimate (MLE) calculation for exponential-trawl processes with general Lévy basis and demonstrate its correctness using several examples.

4.1 MLE Calculation based on Filtering

Recall that the evaluation of the log-likelihood (6) requires the calculations of the conditional intensities $\lambda_{t-}^{(y)}$ and their integrals

$$\int_{t \in (0, T]} \lambda_{t-}^{(y)} dt = v(y) T + \phi \sum_{\mathbf{j}} j_{-y} \int_{t \in (0, T]} p_{t-, t-}(\mathbf{j}) dt, \quad (15)$$

which follows from (7) and $\mathbb{E} \left(C_{t-}^{(-y)} \middle| \mathcal{F}_{t-} \right) = \sum_{\mathbf{j}} j_{-y} p_{t-, t-}(\mathbf{j})$.

However, we do not know the integral $\int_{t \in (0, T]} p_{t-, t-}(\mathbf{j}) dt$ analytically, as the denominator in (9) also depends on t . Hence, we have to calculate (9) in a dense grid of time points—separated by a time gap $\delta_{\text{inactivity}}$ during those inactivity periods—and approximate (15) by linear interpolation. Clearly, the smaller the time gap $\delta_{\text{inactivity}}$, the smaller the numerical error in (15) but the larger the computational burden.

Example 4. Using the true parameters in (11) and simulating a 10-day-long data with $T = 756,000$ (sec.), Fig. 4 shows how an inappropriate choice of $\delta_{\text{inactivity}}$ will depict a wrong log-likelihood surface no matter how long the correct simulated data we supply, where the comparison is made with respect to the first day portion (75,600 (sec.)) of the 10-day-long simulated data. Using the same one-day-long data, Fig. 5 also shows the corresponding log-likelihood function over v^+ or v^- with other parameters fixed at the truth. Including the bottom left panel of Fig. 4, all of the MLE's (solid lines) are reasonably close to the true values (dashed lines) and the likelihood ratio tests suggest that p -values are all greater than 20%.

4.2 Complete-Data Likelihood Inference

Even though in general it would be computationally expensive to calculate the MLE by direct filtering, the maximum complete-data likelihood estimate (MCLE) is much simpler. A comprehensive analysis of the complete-data likelihood inference is performed in the following.

Let $N_t^{(y), A}$ and $N_t^{(y), D}$ be the counting process of the temporary arrival of size y events and the departure of old size y events during the period $(0, T]$. Also let

$$N_t^{\text{type}} \triangleq \sum_{y \in \mathbb{Z} \setminus \{0\}} N_t^{(y), \text{type}}, \quad \text{type} = A, D.$$

Theorem 4. *The complete-data log-likelihood function of the exponential-trawl process is (ignoring the constant)*

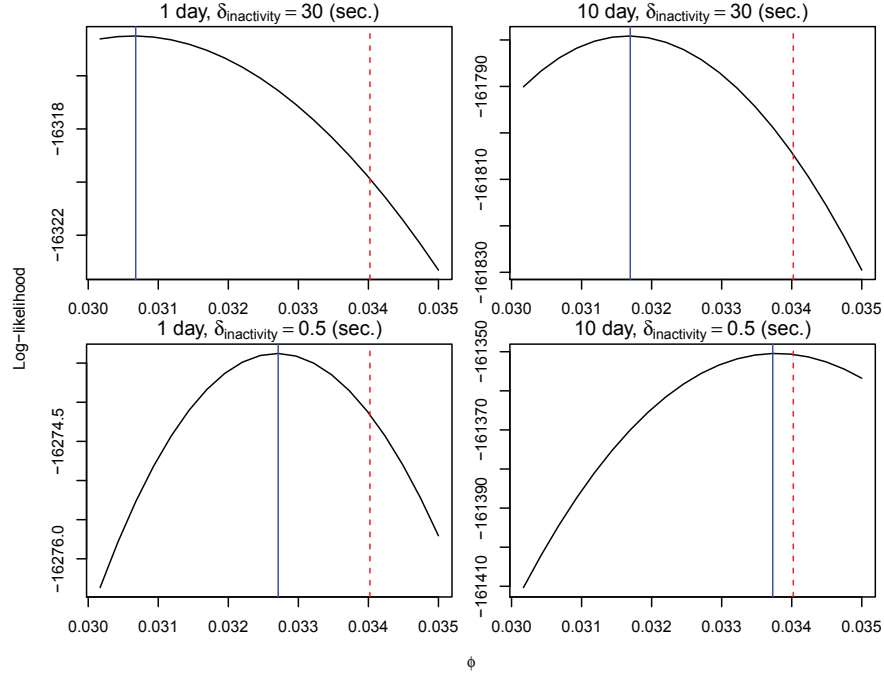


Fig. 4 Log-likelihood plots over ϕ (with v^+ and v^- fixed at the truth) using different $\delta_{\text{inactivity}}$ and a simulated 10-day-long ($T = 756,000$ (sec.)) Skellam exponential-trawl process. The one-day-long data is the first tenth of the simulated data. The dashed lines indicate the true value of ϕ , while the solid lines indicate the optimal value of ϕ in each plot. The p -values using the likelihood ratio test are 0.104% (Top left), 21.0% (Bottom left), 8.82×10^{-13} (Top right) and 46.1% (Bottom right). Code: EPTprocess_MLE.Inference.Simulation.Small-vs.Large.R

$$l_{\mathcal{C}_T}(\theta) = \sum_{y \in \mathbb{Z} \setminus \{0\}} \left(\log(v(y)) \left(N_T^{(y),A} + C_0^{(y)} \right) - v(y) (T + \phi^{-1}) \right) \quad (16)$$

$$+ \log(\phi) (N_T^D - D_0) - \phi \int_{t \in (0,T]} D_t - dt,$$

so the corresponding MCLE's for the Lévy measure and the trawl parameter are

$$\hat{v}_{\text{MCLE}}(y) = \frac{N_T^{(y),A} + C_0^{(y)}}{T + \hat{\phi}_{\text{MCLE}}^{-1}}, \quad y \in \mathbb{Z} \setminus \{0\}, \quad (17)$$

$$\hat{\phi}_{\text{MCLE}} = \frac{\Xi_T + \sqrt{\Xi_T^2 + 4 \frac{N_T^A + N_T^D}{T} \int_{t \in (0,T]} D_t - dt}}{2 \int_{t \in (0,T]} D_t - dt},$$

$$\Xi_T \triangleq N_T^D - D_0 - \frac{1}{T} \int_{t \in (0,T]} D_t - dt.$$

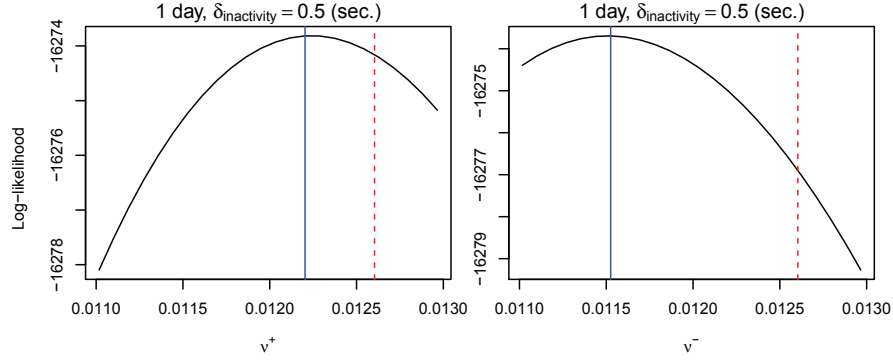


Fig. 5 Log-likelihood plots over either v^+ or v^- for one simulated Skellam exponential-trawl process. The dashed lines indicate the true value, while the solid lines indicate the optimal value of v^+ or v^- in the individual plot. The p -values using the likelihood ratio test are 40.5% (Left) and 33.4% (Right). Code: EPTprocess_MLE_Inference_Simulation_Small_vs_Large.R

Furthermore, the MCLE's above are strong consistent: with probability 1, as $T \rightarrow \infty$,

$$\hat{\phi}_{\text{MCLE}} \rightarrow \phi \text{ and } \hat{v}_{\text{MCLE}}(y) \rightarrow v(y), \quad y \in \mathbb{Z} \setminus \{0\}.$$

We note that $\hat{\phi}_{\text{MCLE}}$ depends on $\int_{t \in (0, T]} D_t -$, the total number of possible departures, weighed by time, at risk during the period $(0, T]$.

4.3 MLE Calculation based on EM Algorithm

In this Subsection, we introduce an EM algorithm that is particularly suitable for exponential-trawl processes, as there are no discretization errors. The EM algorithm is also computationally efficient. Compared with generic optimization methods like *limited-memory BFGS* (*L-BFGS*), the updating scheme suggested by EM can converge to the MLE in a fewer steps and with no error. Clearly, the use of EM needs some extra computations in each step for backward smoothing, but in aggregate EM performs much faster than L-BFGS as EM skips those intermediate filtering calculations during those inactivity periods.

E-Step The linear form of the complete-data log-likelihood (16) allows us to easily take expectation on it with respect to $\mathbb{P}(\cdot | \mathcal{F}_T)$ (under a set of old estimated parameters $\hat{\theta}_{\text{old}}$), which then requires the calculations of the following quantities using the smoothing distribution $p_{t, T}$:

Table 1 The MLE calculations on one simulated Skellam exponential-trawl process using L-BFGS-B procedure in R (with default settings) and EM algorithm (with uniform tolerance 10^{-6} on the parameter space). The R elapsed time is 137.4 (sec.) for L-BFGS-B and 3.3 (sec.) for EM, which is about 40 times speed up. Code: `EPTprocess.MLE.Inference.Simulation.LBFGS.vs.EM.R`

Estimation	Parameter			Log-likelihood	
	v^+	v^-	ϕ	$\delta_{\text{inactivity}} = 0.5$	$\delta_{\text{inactivity}} = 0.01$
Truth	0.01260	0.01111	0.03402	-15,974.98	-15,974.9543
L-BFGS-B	0.01201	0.01128	0.03362	-15,973.92	-15,973.8915
EM	0.01199	0.01126	0.03354	-15,973.91	-15,973.8881

$$\begin{aligned}
\mathbb{E}\left(N_T^{(y),A} \middle| \mathcal{F}_T\right) &= \sum_{0 < t \leq T} \mathbb{P}\left(\Delta C_t^{(y)} = 1 \middle| \mathcal{F}_T\right), \\
\mathbb{E}\left(N_T^{(y),D} \middle| \mathcal{F}_T\right) &= \sum_{0 < t \leq T} \mathbb{P}\left(\Delta C_t^{(y)} = -1 \middle| \mathcal{F}_T\right), \\
\mathbb{E}\left(C_0^{(y)} \middle| \mathcal{F}_T\right) &= \sum_{\mathbf{j}} j_y p_{0,T}(\mathbf{j}), \quad \mathbb{E}(D_0 | \mathcal{F}_T) = \sum_{\mathbf{j}} \|\mathbf{j}\|_1 p_{0,T}(\mathbf{j}), \\
\mathbb{E}(D_{t-} | \mathcal{F}_T) &= \sum_{\mathbf{j}} \|\mathbf{j}\|_1 p_{t-,T}(\mathbf{j}),
\end{aligned} \tag{18}$$

where (13) and (14) will be extensively used. Note that $\mathbb{E}(D_{t-} | \mathcal{F}_T)$ will be a step function of t , so the calculation of $\int_{t \in (0,T]} \mathbb{E}(D_{t-} | \mathcal{F}_T) dt$ is trivially exact.

M-Step Since the *E-Step* generates a Q function that takes the same functional form of θ as (16), the solution to *M-Step* takes the same form as the MCLE in (17), where we just replace each of the hidden data related terms by their smoothing expectations in (18). This can be also viewed as a representation of plug-in principle for (17), i.e., replacing those unknown quantities (e.g. $1_{\{\Delta C_t^{(y)}=1\}}$) by

the known ones (e.g. $\mathbb{P}(\Delta C_t^{(y)} = 1 | \mathcal{F}_T)$). We further use the solution of this *M-Step* for next iteration.

Example 5 (Continued from Example 2).

Using the same simulated Skellam exponential-trawl process path, Table 1 compares the MLE derived from (i) the L-BFGS-B procedure in the `optim` function of the R language (using the default tolerance settings) with that from (ii) the EM algorithm (using the same initial parameter value), which stops until each parameter differs less than a uniform tolerance 10^{-6} .

As expected, using the EM algorithm gives estimation values that are very close to the direct optimization of the log-likelihood function (using $\delta_{\text{inactivity}} = 0.5$). An interesting feature here is that the MLE found by the EM algorithm has a slightly larger log-likelihood value (even for $\delta_{\text{inactivity}} = 0.01$) than by the L-BFGS-B, which might attribute to the numerical insufficiency of the default optimization tolerance setting of R.

The L-BFGS-B procedure uses 27 evaluations of the filtering procedure (9 of them for objective function evaluations and 18 of them for numerical gradients); as a comparison, the EM algorithm takes 12 evaluations of the filtering procedure plus 12 more of the smoothing procedure. In aggregate, the EM algorithm is over 40 times faster than the L-BFGS-B in terms of the computation time.

Starkly different from Example 4, the EM algorithm does not require the fine evaluation of the integrals of $\lambda_{t-}^{(y)}$, so not only the filtering procedure in each iteration of the EM is faster (as it skips the grid calculations of $\lambda_{t-}^{(y)}$ during those inactivity periods) but also the convergent result of EM will maximize the *numerically errorless* log-likelihood (as it has nothing to do with $\delta_{\text{inactivity}}$ to conduct EM). As a conclusion, using EM algorithm to search the MLE for exponential-trawl processes will dominate the direct optimization of log-likelihood both on the numerical quality and on the computation speed.

4.4 Likelihood Inference without the Initial Information

If we consider the complete-data log-likelihood given the information \mathbf{C}_0 , i.e. $l_{\mathcal{C}_T|\mathbf{C}_0}(\theta)$, then the MCLE's are even simpler:

$$\hat{v}_{\text{MCLE}}(y) = \frac{N_T^{(y),A}}{T}, \quad \hat{\phi}_{\text{MCLE}} = \frac{N_T^D}{\int_{t \in (0,T]} D_{t-} dt}.$$

Note that these estimates are the most natural frequency estimates providing that we know the hidden state process \mathbf{C}_t : $v(y)$ is estimated by the sample intensity of all the arrivals of size y events, while ϕ^{-1} is estimated by the average lifetime among all the departures of the temporary events, for the lifetime of any temporary event is exponentially distributed with mean $1/\phi$.

However, here is a subtle statistical inconsistency if one wants to build an EM algorithm based on $l_{\mathcal{C}_T|\mathbf{C}_0}(\theta)$. In practice, all the initial values $\mathbf{C}_0^{(y)}$'s are unknown, so the only way we can work on $l_{\mathcal{C}_T|\mathbf{C}_0}(\theta)$ is to treat them as nuisance parameters. Thus, the EM Q function is defined by

$$Q(\theta', \mathbf{C}_0' | \theta, \mathbf{C}_0) = \mathbb{E}_{\theta} \left(l_{\mathcal{C}_T|\mathbf{C}_0'}(\theta') \middle| \mathbf{C}_0, \mathcal{F}_T \right),$$

which not only requires the smoothing scheme based on $\mathbb{P}_{\theta}(\cdot | \mathcal{F}_T, \mathbf{C}_0)$ —not $\mathbb{P}_{\theta}(\cdot | \mathcal{F}_T)$ —but also finally gives us the MLE of the *joint* log-likelihood function $l_{\mathcal{F}_T}(\theta, \mathbf{C}_0)$ —not the MLE of $l_{\mathcal{F}_T}(\theta)$ nor of $l_{\mathcal{F}_T|Y_0}(\theta)$. On the other hand, one might also define the EM Q function as

$$Q(\theta' | \theta) = \mathbb{E}_{\theta} \left(l_{\mathcal{C}_T|\mathbf{C}_0}(\theta') \middle| \mathcal{F}_T \right),$$

but in this case

$$Q(\theta|\theta) = l_{\mathcal{F}_T|Y_0}(\theta) - \mathbb{E}_\theta(l_{C_0|Y_0}(\theta)|\mathcal{F}_T) \neq l_{\mathcal{F}_T|Y_0}(\theta),$$

which then *breaks* the fundamental monotonicity that guarantees the availability of EM:

$$l_{\mathcal{F}_T|Y_0}(\theta^*) \geq Q(\theta^*|\theta) = \max_{\theta'} Q(\theta'|\theta) \geq Q(\theta|\theta) = l_{\mathcal{F}_T|Y_0}(\theta).$$

Therefore, even though the direct filtering allows the calculations of the MLE whenever we include the initial information Y_0 or not (i.e. to maximize $l_{\mathcal{F}_T}(\theta)$ or $l_{\mathcal{F}_T|Y_0}(\theta)$), a *correct* EM-based inference will automatically enforce the consideration of Y_0 (i.e. to maximize $l_{\mathcal{F}_T}(\theta)$ using EM). This is a bit different from likelihood inference for marked point processes, which usually ignores the effect of the initial value Y_0 . This mild difference will clearly disappear asymptotically as $T \rightarrow \infty$, but here we still prefer to present a complete likelihood analysis for trawl processes instead of treating them the same as marked point processes.

5 Likelihood Inference for Non-negative Exponential-Trawl Processes

In this Section, we focus on exponential-trawl processes that are always non-negative. Then all the negative movements of this type of processes must attribute to the departures of the positive events in the trawl, so it is natural to split up Y into the counting process of size y jumps

$$N_t^{(y)} \triangleq \sum_{0 < s \leq t} 1_{\{\Delta Y_s = y\}}, \quad y \in \mathbb{Z} \setminus \{0\},$$

which relates to $C_t^{(y)}$ via

$$C_t^{(y)} = C_0^{(y)} + N_t^{(y)} - N_t^{(-y)}. \quad (19)$$

Then, as mentioned in the end of Subsection 2.2,

$$Y_t = \sum_{y=1}^{\infty} y C_t^{(y)} = \sum_{y=1}^{\infty} y C_0^{(y)} + \sum_{y=1}^{\infty} y (N_t^{(y)} - N_t^{(-y)}).$$

Clearly, the path of Y_t reveals the path of each of the individual $N_t^{(y)}$ for $y \in \mathbb{Z} \setminus \{0\}$, so $N_t^{(y)} \in \mathcal{F}_t$. Thus, the only unknown objects here are $C_0^{(y)}$'s, for we just see $Y_0 = \sum_{y=1}^{\infty} y C_0^{(y)}$ and all the departures resulted from $C_0^{(y)}$'s. If we can know $C_0^{(y)}$, then we will see the complete path of $C_t^{(y)}$ and hence likelihood inference will be particularly tractable.

5.1 Partial Likelihood Inference

We can specialize Corollary 1 using (8) and write down the log-likelihood for the non-negative case (ignoring the constant):

$$\begin{aligned} l_{\mathcal{F}_T}(\theta) &= \sum_{y=1}^{\infty} \left(\log(v(y)) N_T^{(y)} - v(y) T \right) \\ &\quad + \log(\phi) N_T^{(-)} - \phi \int_{t \in (0, T]} \mathbb{E}_{\theta} \left(C_{t-}^{(+)} \middle| \mathcal{F}_{t-} \right) dt \\ &\quad + \sum_{0 < t \leq T} \sum_{y=1}^{\infty} \log \mathbb{E}_{\theta} \left(C_{t-}^{(y)} \middle| \mathcal{F}_{t-} \right) 1_{\{\Delta Y_t = -y\}} + l_{Y_0}(\theta), \end{aligned}$$

where $N_T^{(-)} \triangleq \sum_{y=1}^{\infty} N_T^{(-y)}$ and $C_{t-}^{(+)} \triangleq \sum_{y=1}^{\infty} C_{t-}^{(y)}$.

Like the general case we studied in Section 4, there are no analytic expressions available for the filtering expectations $\mathbb{E}_{\theta} \left(C_{t-}^{(y)} \middle| \mathcal{F}_{t-} \right)$ and the initial likelihood $l_{Y_0}(\theta)$, so finding $\hat{\theta}_{\text{MLE}}$ also requires the EM techniques we introduced before. However, the first part of $l_{\mathcal{F}_T}(\theta)$ that involves $v(y)$'s is particularly analytically tractable, so this leads us to consider the following maximum partial likelihood estimate (MPLE) for the Lévy measure:

$$\hat{v}_{\text{MPLE}}(y) = \frac{N_T^{(y)}}{T}, \quad y = 1, 2, 3, \dots,$$

which is a non-parametric moment estimate that is apparent from the non-negative setting.

Even though \hat{v}_{MPLE} is not \hat{v}_{MLE} , it has several advantages. First, it has strong consistency, i.e., with probability 1, $\hat{v}_{\text{MPLE}}(y) \rightarrow v(y)$ as $T \rightarrow \infty$. Second, it is asymptotically equivalent to the MCLE, because

$$\hat{v}_{\text{MCLE}}(y) = \frac{N_T^{(y)} + C_0^{(y)}}{T + \hat{\phi}_{\text{MCLE}}^{-1}} = \frac{\frac{N_T^{(y)}}{T} + \frac{C_0^{(y)}}{T}}{1 + \frac{\hat{\phi}_{\text{MCLE}}^{-1}}{T}} \approx \hat{v}_{\text{MPLE}},$$

where the MCLE of θ is simply given from (17) but we need to replace those D_{t-} by $C_{t-}^{(+)}$. Third, it allows to estimate each component of the Lévy measure separately from themselves and from ϕ , as given a long enough path of Y , including the initial value $C_0^{(y)}$ and $\hat{\phi}_{\text{MCLE}}^{-1}$ has no strong improvement on the estimation quality of \hat{v}_{MPLE} .

Alternatively, a parameterized common intensity function $v(y|\eta)$ can be used, where η is some finite dimensional parameter. Then the MPLE is found by solving

$$\hat{\eta}_{\text{MPLE}} \triangleq \underset{\eta}{\operatorname{argmax}} \sum_{y=1}^{\infty} \left(\log(v(y|\eta)) N_T^{(y)} - v(y|\eta) T \right)$$

and letting $\hat{v}_{\text{MPLE}}(y) = v(y|\hat{\eta}_{\text{MPLE}})$.

To infer on the trawl parameter ϕ , we can simply plug-in the \hat{v}_{MPLE} (either parametric or non-parametric) and then do the filtering procedure to calculate $\mathbb{E}_{\theta=(\hat{v}_{\text{MPLE}}, \phi)} \left(C_{t-}^{(y)} \middle| \mathcal{F}_{t-} \right)$ for $y = 1, 2, \dots$ and $t \in (0, T]$. Combining this with an (one-dimensional) optimization procedure we can find

$$\hat{\phi}_{\text{MPLE}} \triangleq \underset{\phi}{\operatorname{argmax}} l_{\mathcal{F}_T}(\hat{v}_{\text{MPLE}}, \phi).$$

5.2 Estimate the Missing Initial Missing Values

Except the Poisson case ($Y_0 = C_0^{(1)}$, $C_0^{(y)} = 0$ for all $y > 1$ and hence in particular $\hat{\theta}_{\text{MCLE}} = \hat{\theta}_{\text{MLE}}$), every $C_0^{(y)}$'s are missing, so in principle we need to estimate these initial values in order to get (an approximation of) $\hat{\theta}_{\text{MCLE}}$. Indeed, the EM algorithm also does so through the smoothing expectations

$$\mathbb{E} \left(C_{t-}^{(y)} \middle| \mathcal{F}_T \right) = \mathbb{E} \left(C_0^{(y)} \middle| \mathcal{F}_T \right) + N_t^{(y)} - N_t^{(-y)},$$

but it just iterates (17) until converges. Nevertheless, there is another simpler estimation of $C_0^{(y)}$ thanks to the special non-negative feature.

The following Proposition only relies on the fact that Y_t is non-negative and in fact does not depend on the choice of the trawl.

Proposition 1. *Assume that the (general) trawl process Y_t is non-negative. If*

$$C_{0,T}^{(y),\text{L}} \triangleq \sup_{t \in [0, T]} \left(N_t^{(-y)} - N_t^{(y)} \right), \quad C_{0,T}^{(y),\text{U}} \triangleq \left\lfloor \frac{Y_0 - \sum_{y' \neq y} y' C_{0,T}^{(y'),\text{L}}}{y} \right\rfloor,$$

where $N_0^{(y)} \triangleq 0$ conventionally and $\lfloor x \rfloor$ means the integer part of x , then

$$C_{0,T}^{(y),\text{U}} \geq C_0^{(y)} \geq C_{0,T}^{(y),\text{L}}.$$

Furthermore,

$$\lim_{T \rightarrow \infty} C_{0,T}^{(y),\text{U}} = \lim_{T \rightarrow \infty} C_{0,T}^{(y),\text{L}} = C_0^{(y)}.$$

Thus, a straightforward and sharp estimation to $C_0^{(y)}$ can be given by, e.g.,

$$\hat{C}_0^{(y)} \triangleq \left\lfloor \frac{C_{0,T}^{(y),U} + C_{0,T}^{(y),L}}{2} \right\rfloor,$$

so use this estimation in (17) will give an estimate of θ that is almost as good as $\hat{\theta}_{\text{MCLE}}$.

Example 6. Figure 6 illustrates Proposition 1 with a non-negative geometric Lévy basis, where

$$\begin{aligned} v(y|\eta) &= \|v\| \eta (1-\eta)^{y-1}, \quad y = 1, 2, \dots, \\ \|v\| &= 3, \quad \eta = 0.5, \quad \phi = 0.5, \quad T = 100. \end{aligned}$$

The paths of the upper bound $C_{0,t}^{(y),U}$ and the lower bound $C_{0,t}^{(y),L}$ are shown as step functions of time t in Fig. 6. We can observe a strong convergent pattern, as all the bounds for different y converge after $t > 15$ —the perfect estimations of the initial values $C_0^{(y)}$'s. Furthermore, as $Y_0 = C_0^{(1)} + 2C_0^{(2)} + 4C_0^{(4)}$ in this case, all the other $C_0^{(y)}$'s for $y \neq 1, 2, 4$ must be zero. We then have discovered all the initial values and can use them to conduct MCLE by (17).

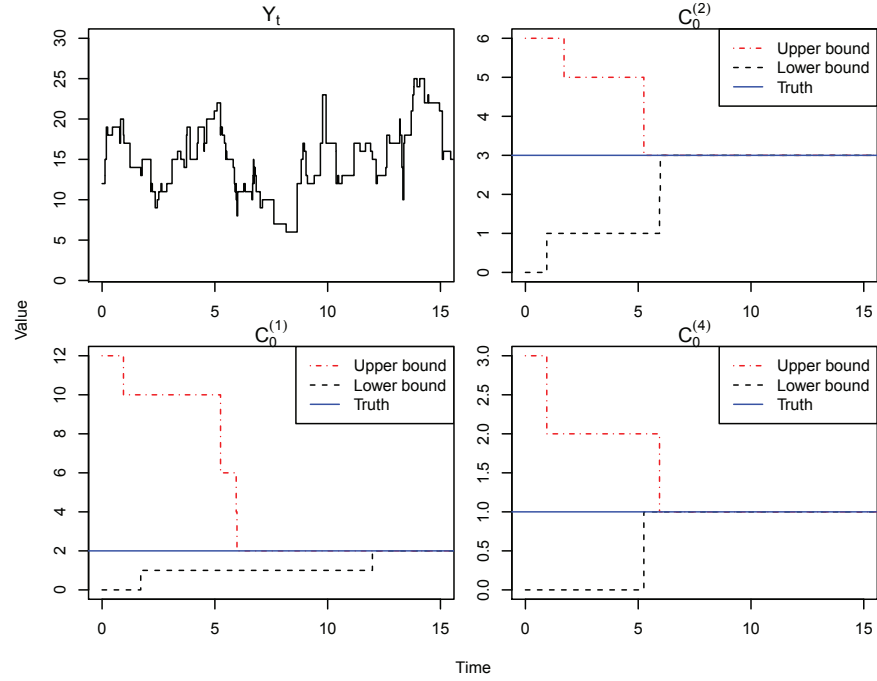


Fig. 6 *Top left:* A simulated path for the exponential-trawl process Y_t using non-negative geometric Lévy basis. *Top right, Bottom left, Bottom right:* Paths of $C_{0,t}^{(y),U}$ and $C_{0,t}^{(y),L}$ along with the true $C_0^{(y)}$ for $y = 1, 2, 4$. Code: `EPTprocess.NonNegativeInitialEstimate.R`

6 Conclusion

In this Chapter, we studied likelihood-based inference of the trawl processes by explicitly working on the filtering and smoothing procedures inherited from this model. It is plausible and practically implementable under the exponential trawl. We used some simulation examples to justify the correctness of our procedures.

The major contribution of this Chapter is to provide an easiest beginning step toward likelihood inference for all of the other more general trawl processes, which might even allow the inclusion of a non-stationary Lévy process component. [21] calls it a fleeting price process and extensively uses it for the study of high frequency financial econometrics.

The filters for the fleeting price process they proposed will allow an econometrically interesting decomposition of observed prices into equilibrium prices and market microstructure noises. More empirical analysis about these will be addressed in the future work.

Appendix: Proofs and Derivations

6.1 Heuristic Proof of Theorem 1

Our heuristic derivation starts from the following prediction decomposition of the Radon-Nikodym derivative:

$$\log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{\mathcal{F}_T^X | X_0} = \int_{t \in (0, T]} \log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{X_t | \mathcal{F}_{t-}^X}, \quad (20)$$

where the integral over $t \in (0, T]$ means a continuous sum of the integrand random variables. Thus,

$$\begin{aligned} \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{X_t | \mathcal{F}_{t-}^X} &= \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{\Delta X_t | \mathcal{F}_{t-}^X} \\ &= \sum_{y \in \mathbb{Z} \setminus \{0\}} \frac{\mathbb{P}(\Delta X_t = y | \mathcal{F}_{t-}^X)}{\mathbb{Q}(\Delta X_t = y | \mathcal{F}_{t-}^X)} 1_{\{\Delta X_t = y\}} + \frac{\mathbb{P}(\Delta X_t = 0 | \mathcal{F}_{t-}^X)}{\mathbb{Q}(\Delta X_t = 0 | \mathcal{F}_{t-}^X)} 1_{\{\Delta X_t = 0\}} \\ &= \sum_{y \in \mathbb{Z} \setminus \{0\}} \frac{\lambda_{t-}^{(y), \mathbb{P}} dt}{\lambda_{t-}^{(y), \mathbb{Q}} dt} 1_{\{\Delta X_t = y\}} + \frac{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{P}} dt}{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{Q}} dt} 1_{\{\Delta X_t = 0\}}, \end{aligned}$$

where the first equality follows because X_{t-} is known in \mathcal{F}_{t-}^X ; the third equality follows from (5). Therefore, (20) can be rewritten as

$$\begin{aligned}
\int_{t \in (0, T]} \log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{X_t | \mathcal{F}_{t-}^X} &= \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \left(\frac{\lambda_{t-}^{(y), \mathbb{P}} dt}{\lambda_{t-}^{(y), \mathbb{Q}} dt} \right) 1_{\{\Delta X_t = y\}} \\
&\quad + \int_{\{t \in (0, T] : \Delta X_t = 0\}} \log \left(\frac{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{P}} dt}{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{Q}} dt} \right) \\
&= \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \left(\frac{\lambda_{t-}^{(y), \mathbb{P}}}{\lambda_{t-}^{(y), \mathbb{Q}}} \right) 1_{\{\Delta X_t = y\}} \\
&\quad - \int_{t \in (0, T]} \sum_{y \in \mathbb{Z} \setminus \{0\}} \left(\lambda_{t-}^{(y), \mathbb{P}} - \lambda_{t-}^{(y), \mathbb{Q}} \right) dt,
\end{aligned}$$

where the second equality follows from $\log(1-x) \approx -x$ for small x and $\{t \in (0, T] : \Delta X_t \neq 0\}$ has Lebesgue measure 0.

6.2 Heuristic Proof of Theorem 2

6.2.1 Update by inactivity

We want to update $p_{\tau, \tau}(\mathbf{j})$ by incorporating the information $\mathcal{F}_{(\tau, t)} \triangleq \sigma(\{\Delta Y_s = 0, \tau < s < t\})$ using Bayes' Theorem:

$$\begin{aligned}
\mathbb{P}(\mathbf{C}_{t-} = \mathbf{j} | \mathcal{F}_{t-}) &= \mathbb{P}(\mathbf{C}_{\tau} = \mathbf{j} | \mathcal{F}_{t-}) = \mathbb{P}(\mathbf{C}_{\tau} = \mathbf{j} | \mathcal{F}_{\tau}, \mathcal{F}_{(\tau, t)}) \\
&\propto \mathbb{P}(\mathcal{F}_{(\tau, t)} | \mathcal{F}_{\tau}, \mathbf{C}_{\tau} = \mathbf{j}) \mathbb{P}(\mathbf{C}_{\tau} = \mathbf{j} | \mathcal{F}_{\tau}),
\end{aligned}$$

where the first equality holds because there is no activity of Y_s for $s \in (\tau, t)$ and hence the hidden state \mathbf{C} must stay the same.

Using the prediction decomposition, we have

$$\begin{aligned}
\log \mathbb{P}(\mathcal{F}_{(\tau, t)} | \mathcal{F}_{\tau}, \mathbf{C}_{\tau} = \mathbf{j}) &= \int_{s \in (\tau, t)} \log \mathbb{P}(\Delta Y_s = 0 | \mathcal{F}_{\tau}, \mathcal{F}_{(\tau, s)}, \mathbf{C}_{\tau} = \mathbf{j}) \\
&= \int_{s \in (\tau, t)} \log \left(1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y) ds - \sum_{y \in \mathbb{Z} \setminus \{0\}} \phi_j y ds \right) \\
&= - \sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y) (t - \tau) - \phi \|\mathbf{j}\|_1 (t - \tau),
\end{aligned}$$

where the second equality intuitively holds because we know the instantaneous departure probability of a size y event at time s is $\phi C_{s-}^{(y)} ds$ but $C_{s-}^{(y)} = C_{\tau}^{(y)} = j_y$ under $\mathcal{F}_{(\tau, s)}$; the third equality follows from $\log(1-x) \approx -x$ for small x . Therefore,

$$\mathbb{P}(\mathbf{C}_{t-} = \mathbf{j} | \mathcal{F}_{t-}) \propto e^{-\phi \|\mathbf{j}\|_1 (t - \tau)} \mathbb{P}(\mathbf{C}_{\tau} = \mathbf{j} | \mathcal{F}_{\tau}),$$

where we throw out the term $\exp(-\sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y)(t - \tau))$ because it doesn't depend on \mathbf{j} . Normalizing the equation above leads to the desired result.

6.2.2 Update by jump

We want to update $p_{\tau-, \tau-}(\mathbf{j})$ by incorporating the piece of information, $\Delta Y_\tau = y$. First note that

$$\begin{aligned} \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_\tau) &= \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_{\tau-}, \Delta Y_\tau = y) \\ &= \mathbb{P}\left(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} \middle| \mathcal{F}_{\tau-}, \Delta Y_\tau = y\right) \\ &\quad + \mathbb{P}\left(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} + \mathbf{1}^{(-y)} \middle| \mathcal{F}_{\tau-}, \Delta Y_\tau = y\right), \end{aligned}$$

which corresponds to the arrival of a new size y event and the departure of an old size $-y$ event.

For the first term,

$$\begin{aligned} &\mathbb{P}\left(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} \middle| \mathcal{F}_{\tau-}, \Delta Y_\tau = y\right) \\ &= \frac{\mathbb{P}\left(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)}, \Delta Y_\tau = y \middle| \mathcal{F}_{\tau-}\right)}{\mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})} \\ &= \frac{\mathbb{P}\left(\mathbf{C}_\tau = \mathbf{j}, \Delta Y_\tau = y \middle| \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)}, \mathcal{F}_{\tau-}\right) \mathbb{P}\left(\mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} \middle| \mathcal{F}_{\tau-}\right)}{\mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})} \\ &= \frac{\mathbb{P}\left(\Delta \mathbf{C}_\tau = \mathbf{1}^{(y)} \middle| \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)}, \mathcal{F}_{\tau-}\right) \mathbb{P}\left(\mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} \middle| \mathcal{F}_{\tau-}\right)}{\mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})} \\ &= \frac{\nu(y)}{\lambda_{\tau-}^{(y)}} \mathbb{P}\left(\mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} \middle| \mathcal{F}_{\tau-}\right), \end{aligned}$$

where the fourth equality follows from (3) (using $\mathcal{C}_{\tau-} \supseteq \mathcal{F}_{\tau-}$) and (5).

Using similar arguments, the second term is

$$\mathbb{P}\left(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} + \mathbf{1}^{(-y)} \middle| \mathcal{F}_{\tau-}, \Delta Y_\tau = y\right) = \frac{\phi(j-y+1)}{\lambda_{\tau-}^{(y)}} \mathbb{P}\left(\mathbf{C}_{\tau-} = \mathbf{j} + \mathbf{1}^{(-y)} \middle| \mathcal{F}_{\tau-}\right).$$

Combining all of these gives us the required result.

6.3 Heuristic Proof of Theorem 3

The case of updating smoothing distribution $p_{\tau-,T}(\mathbf{j})$ due to inactivity is trivial because the hidden configuration \mathbf{C} must stay unchanged because of the inactivity during the time period $[t, \tau)$.

6.3.1 Update by jump

We now consider the case of (backward) updating the smoothing distribution $p_{\tau,T}(\mathbf{j})$ due to the jump $\Delta Y_\tau = y$. Then

$$\begin{aligned} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T) &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j}, \mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)} | \mathcal{F}_T) + \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j}, \mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)} | \mathcal{F}_T) \\ &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)}) \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)} | \mathcal{F}_T) \\ &\quad + \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)}) \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)} | \mathcal{F}_T). \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{k}) &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}) \tag{21} \\ &= \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_\tau) \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau)}{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathcal{F}_\tau)} \\ &\quad \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_\tau) \mathbb{P}(\Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-})}{\times \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_{\tau-})} \\ &= \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathcal{F}_\tau) \mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})}{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k}, \Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_{\tau-})} \\ &= \frac{\lambda_{\tau-}^{(y)} dt}{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathcal{F}_\tau)}, \end{aligned}$$

where the first equality holds due to the Markov property of \mathbf{C}_t , a heuristic derivation is given later; the second and third equalities follow from the Bayes' Theorem. Since

$$\begin{aligned} \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)}, \Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) &= \mathbb{P}(\Delta \mathbf{C}_\tau = \mathbf{1}^{(y)} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) \\ &= v(y) dt, \\ \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)}, \Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) &= \mathbb{P}(\Delta \mathbf{C}_\tau = -\mathbf{1}^{(-y)} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) \\ &= \phi j_{-y} dt, \end{aligned}$$

combining all of these gives us the required result.

6.3.2 Derivation of (21)

Let $\mathcal{F}_{(\tau, T]} \triangleq \sigma(\{Y_t\}_{\tau < t \leq T})$ and $\mathcal{C}_{(\tau, T]} \triangleq \sigma(\{C_t\}_{\tau < t \leq T})$. Note that heuristically the Bayes' Theorem implies

$$\begin{aligned} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{k}) &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathcal{F}_{(\tau, T]}, \mathbf{C}_\tau = \mathbf{k}) \\ &= \frac{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}, \mathbf{C}_{\tau-} = \mathbf{j}}}{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}}} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}). \end{aligned}$$

Since $\mathcal{F}_{(\tau, T]} \subseteq \mathcal{C}_{(\tau, T]}$ (each $Y_t = \sum_{y \in \mathbb{Z} \setminus \{0\}} C_t^{(y)}$), the Markov property of \mathbf{C}_t implies

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}, \mathbf{C}_{\tau-} = \mathbf{j}} = \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}},$$

because given the current information \mathbf{C}_τ the information in the past $\mathbf{C}_{\tau-}$ is irrelevant. This then proves that

$$\mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{k}) = \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}).$$

6.4 Proof of Theorem 4

Since each $C_t^{(y)}$ is independent for different y , the complete-data log-likelihood can be written as

$$l_{\mathcal{C}_T}(\theta) = \sum_{y \in \mathbb{Z} \setminus \{0\}} l_{\mathcal{C}_T^{(y)} | C_0^{(y)}}(\theta) + \sum_{y \in \mathbb{Z} \setminus \{0\}} l_{C_0^{(y)}}(\theta),$$

where we recall that $\mathcal{C}_t^{(y)}$ is the natural filtration generated by $C_t^{(y)}$,

$$\begin{aligned} l_{\mathcal{C}_T^{(y)} | C_0^{(y)}}(\theta) &= \sum_{0 < t \leq T} \left(\log(v(y)) 1_{\{\Delta C_t^{(y)} = 1\}} + \log(\phi C_{t-}^{(y)}) 1_{\{\Delta C_t^{(y)} = -1\}} \right) \\ &\quad - \int_{t \in (0, T]} \left(v(y) + \phi C_{t-}^{(y)} \right) dt \\ &= \log(v(y)) N_T^{(y), A} - v(y) T + \log(\phi) N_T^{(y), D} - \phi \int_{t \in (0, T]} C_{t-}^{(y)} dt, \end{aligned}$$

where the first equality follows directly from Theorem 1 (ignoring the constant), and

$$l_{C_0^{(y)}}(\theta) = C_0^{(y)} (\log v(y) - \log \phi) - \frac{v(y)}{\phi}$$

because of $C_0^{(y)} \sim \text{Poisson}(v(y)/\phi)$. Thus, collecting terms will give us the required result (16). The derivations of the MCLE are elementary.

Let

$$\|v\| \triangleq \int v(dy) = \sum_{y=1}^{\infty} v(y).$$

The ergodicity of D_{t-} implies that as $T \rightarrow \infty$

$$\frac{1}{T} \int_{t \in (0, T]} D_{t-} dt \rightarrow \mathbb{E}(D_{t-}) = \frac{\|v\|}{\phi}.$$

Since $\frac{N_T^D}{T} \approx \frac{N_T^A}{T} \rightarrow \|v\|$, we have

$$\frac{\Xi_T}{T} = \frac{N_T^D}{T} - \frac{D_0 + T^{-1} \int_{t \in (0, T]} D_{t-} dt}{T} \rightarrow \|v\|, \text{ too.}$$

Thus,

$$\begin{aligned} \hat{\phi}_{\text{MCLE}} &= \frac{\frac{\Xi_T}{T} + \sqrt{\left(\frac{\Xi_T}{T}\right)^2 + 4T^{-1} \frac{N_T^A + N_T^D}{T} T^{-1} \int_{t \in (0, T]} D_{t-} dt}}{2T^{-1} \int_{t \in (0, T]} D_{t-} dt} \\ &\rightarrow \frac{\|v\| + \sqrt{\|v\|^2 + 0}}{2 \frac{\|v\|}{\phi}} = \phi. \end{aligned}$$

Finally, for any $y \in \mathbb{Z} \setminus \{0\}$, $\frac{N_T^{(y)}}{T} \rightarrow v(y)$ and $\hat{\phi}_{\text{MCLE}}^{-1} \rightarrow \phi^{-1} < \infty$, so we easily have

$$\hat{v}_{\text{MCLE}}(y) = \frac{\frac{N_T^{(y)}}{T} + \frac{C_0^{(y)}}{T}}{1 + \frac{\hat{\phi}_{\text{MCLE}}^{-1}}{T}} \rightarrow v(y) \text{ as well.}$$

6.5 Proof of Proposition 1

As $C_t^{(y)} \geq 0$, (19) implies that

$$C_0^{(y)} \geq C_{0,T}^{(y),L} = \sup_{t \in [0, T]} \left(M_t^{(-y)} - N_t^{(y)} \right), \quad y = 1, 2, \dots,$$

where we set $N_0^{(y)} \triangleq 0$ conventionally. Now

$$C_0^{(y)} = \frac{Y_0 - \sum_{y' \neq y} y' C_0^{(y')}}{y} \leq \left\lfloor \frac{Y_0 - \sum_{y' \neq y} y' C_{0,T}^{(y'),L}}{y} \right\rfloor = C_{0,T}^{(y),U},$$

so we have

$$C_{0,T}^{(y),U} \geq C_0^{(y)} \geq C_{0,T}^{(y),L}.$$

Let $N_t^{(-y),*}$ be the counting process of $-y$ jumps resulted from the departures of those initial events of size y that constitute $C_0^{(y)}$. Let τ be the time when $N^{(-y),*}$ achieve $C_0^{(y)}$. Then we have

$$\begin{aligned} C_{0,T}^{(y),L} &= C_{0,\tau}^{(y),L} \vee \sup_{t \in (\tau, T]} \left(N_t^{(-y),*} - \left(N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right) \right) \right) \\ &= C_{0,\tau}^{(y),L} \vee \left(C_0^{(y)} - \inf_{t \in (\tau, T]} \left(N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right) \right) \right). \end{aligned}$$

Observe that $N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right)$ is a $M/G/\infty$ queue initiated at state 0, so by the ergodicity we must have with probability 1

$$\lim_{T \rightarrow \infty} \inf_{t \in (\tau, T]} \left(N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right) \right) = 0.$$

This then shows that actually

$$\lim_{T \rightarrow \infty} C_{0,T}^{(y),L} = C_{0,\tau}^{(y),L} \vee C_0^{(y)} = C_0^{(y)},$$

where the last equality follows because $C_{0,\tau}^{(y),L} \leq C_0^{(y)}$. Correspondingly,

$$\lim_{T \rightarrow \infty} C_{0,T}^{(y),U} = \left\lfloor \frac{Y_0 - \sum_{y' \neq y} y' C_0^{(y')}}{y} \right\rfloor = C_0^{(y)}.$$

References

1. Asmussen, S.: Applied Probability and Queues. Springer, New York (2003)
2. Barndorff-Nielsen, O.E.: Stationary infinitely divisible processes. *Braz. J. Probab. Stat.* **25**, 294–322 (2011)
3. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Ambit processes and stochastic partial differential equations. In: G. Di Nunno, B. Øksendal (eds.) *Advanced Mathematical Methods for Finance*, pp. 35–74. Springer, Berlin Heidelberg (2011)
4. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. *ArXiv e-prints* (2012). Unpublished paper, Department of Mathematics, Imperial College London
5. Barndorff-Nielsen, O.E., Lunde, A., Shephard, N., Veraart, A.E.D.: Integer-valued trawl processes: A class of stationary infinitely divisible processes. *Scandinavian Journal of Statistics*

- 41, 693–724 (2014)
6. Barndorff-Nielsen, O.E., Pollard, D.G., Shephard, N.: Integer-valued Lévy processes and low latency financial econometrics. *Quantitative Finance* **12**, 587–605 (2012)
7. Barndorff-Nielsen, O.E., Schmiegel, J.: Ambit processes; with applications to turbulence and tumour growth. In: F.E. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal, T. Zhang (eds.) *Stochastic Analysis and Applications*, pp. 93–124. Springer, Berlin Heidelberg (2007)
8. Bartlett, M.S.: *An Introduction to Stochastic Processes, with Special Reference to Methods and Applications*, 3 edn. Cambridge University Press, Cambridge (1978)
9. Cameron, C.A., Trivedi, P.K.: *Regression Analysis of Count Data*. Cambridge University Press, Cambridge (1998)
10. Cui, Y., Lund, R.: A new look at time series of counts. *Biometrika* **96**, 781–792 (2009)
11. Daley, D., Vere-Jones, D.: Evolutionary processes and predictability. In: *An Introduction to the Theory of Point Processes, Probability and Its Applications*, pp. 355–456. Springer, New York (2008)
12. Davis, R.A., Wu, R.: A negative binomial model for time series of counts. *Biometrika* **96**, 735–749 (2009)
13. Fokianos, K., Kedem, B.: Regression theory for categorical time series. *Statistical Science* **18**, 357–376 (2003)
14. Jacobs, P.A., Lewis, P.A.W.: Discrete time series generated by mixtures. I: Correlational and runs properties. *Journal of the Royal Statistical Society. Series B (Methodological)* **40**, 94–105 (1978)
15. Jung, R., Tremayne, A.: Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis* **95**, 59–91 (2011)
16. Lindley, D.V.: The estimation of velocity distributions from counts. In: *Proceedings of the International Congress of Mathematicians*, vol. 3, pp. 427–444. North-Holland, Amsterdam (1956)
17. McKenzie, D.J.: Measuring inequality with asset indicators. *Journal of Population Economics* **18**, 229–260 (2005)
18. Reynolds, J.F.: On the autocorrelation and spectral functions of queues. *Journal of Applied Probability* **5**, 467–475 (1968)
19. Rudemo, M.: State estimation for partially observed Markov chains. *Journal of Mathematical Analysis and Applications* **44**, 581–611 (1973)
20. Rudemo, M.: Prediction and smoothing for partially observed Markov chains. *Journal of Mathematical Analysis and Applications* **49**, 1–23 (1975)
21. Shephard, N., Yang, J.J.: Continuous time analysis of fleeting discrete price moves. *ArXiv e-prints* (2014). Unpublished paper, Department of Statistics, Harvard University
22. Surgailis, D., Rosinski, J., Mandrekar, V., Cambanis, S.: Stable mixed moving averages. *Probability Theory and Related Fields* **97**, 543–558 (1993)
23. Weiß, C.: Thinning operations for modeling time series of counts—a survey. *AStA Advances in Statistical Analysis* **92**, 319–341 (2008)
24. Wolpert, R.L., Brown, L.D.: Stationary infinitely-divisible Markov processes with non-negative integer values (2011). Working paper, Department of Statistics, Duke University
25. Wolpert, R.L., Taqqu, M.S.: Fractional Ornstein-Uhlenbeck Lévy processes and the telecom process: Upstairs and downstairs. *Signal Processing* **85**, 1523–1545 (2005)
26. Zhu, R., Joe, H.: A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models* **19**, 235–254 (2003)